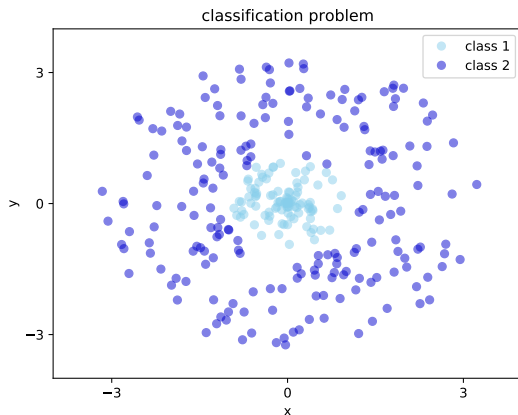


Fondamentaux théoriques du machine learning



Minimization

Exercice 1 : Gradient :

Compute the gradient of $J(\Omega, z)$ with respect to Ω and deduce the minimizer Ω^* .

- ▶ z is fixed
- ▶ we can see Ω has a vector of \mathbb{R}^{Kd} .

It is sufficient to compute the gradient with respect to all the ω_k separately.

Minimization

We compute the gradient of J with respect to ω_{k_0} .

$$\begin{aligned}\nabla_{\omega_{k_0}} J &= \nabla_{\omega_{k_0}} \sum_{i=1}^n \sum_{k=1}^K z_i^k \|x_i - \omega_k\|^2 \\ &= \sum_{i=1}^n \nabla_{\omega_{k_0}} \sum_{k=1}^K z_i^k \|x_i - \omega_k\|^2 \\ &= \sum_{i=1}^n z_i^{k_0} \nabla_{\omega_{k_0}} \|x_i - \omega_{k_0}\|^2\end{aligned}\tag{1}$$

We need to compute

$$\nabla_{\omega_{k_0}} ||x_i - \omega_{k_0}||^2 \quad (2)$$

Let

$$u = \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ y \mapsto ||y||^2 \end{cases}$$

$$v_i = \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ \omega \mapsto x_i - \omega \end{cases}$$

$$w_i = \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ \omega \mapsto ||x_i - \omega||^2 \end{cases}$$

We have

$$w_i = u \circ v_i \quad (3)$$

Differentials

We already now that

$$\forall (y, h) \in (\mathbb{R}^d)^2, du_y(h) = \langle 2y, h \rangle \quad (4)$$

and that

$$\forall (\omega, h) \in (\mathbb{R}^d)^2, d(v_i)_\omega(h) = -h \quad (5)$$

By composition,

$$\nabla_{\omega_{k_0}} \|x_i - \omega_{k_0}\|^2 = -2(x_i - \omega_{k_0}) \quad (6)$$

Gradient cancellation

This gradient cancels if

$$\begin{aligned}\nabla_{\omega_{k_0}} J &= \sum_{i=1}^n z_i^{k_0} \nabla_{\omega_{k_0}} \|x_i - \omega_{k_0}\|^2 \\ &= - \sum_{i=1}^n z_i^{k_0} 2(x_i - \omega_{k_0}) \\ &= 0\end{aligned}\tag{7}$$

Gradient cancellation

This gradient cancels if

$$2\omega_{k_0} \sum_{i=1}^n z_i^{k_0} = 2 \sum_{i=1}^n z_i^{k_0} x_i \quad (8)$$

or equivalently

$$\omega_{k_0} = \frac{\sum_{i=1}^n z_i^{k_0} x_i}{\sum_{i=1}^n z_i^{k_0}} \quad (9)$$

Hence, the minimizer $w_{k_0}^*$ is the average of its cluster.